INFORMATION RETRIEVAL

A Critical View

Edited by GEORGE SCHECTER

GETTING IT OUT OF OUR SYSTEM

THEODOR H. NELSON
Harcourt, Brace & World, Inc.
New York, New York

SUMMARY

Specialist access and muster of information—usually the underlying task of "information retrieval"—may not be best accomplished either by indexing techniques (document retrieval) or queriable information networks (content retrieval). Digital text storage and display make possible the creation of at least one potentially powerful *medium*: the hypertext, or non-linear text system. As a medium, not a facility, it will share traits of book and film. Machines will present complex interconnected text units to users who will weave their way through them at a glowing screen. Successive user choices will evoke successive presentations which have been created—written and structurally designed, and signed—by individuals and teams. Departing completely from the convention of linear text as we know it, the hypertext will probably be much more effective for everything.

Several parallels exist between the (potential) development of the hypertext medium and the actual development of the motion picture in its early years: the inventors of the technology underestimated the importance and generality of the resulting medium; techniques of predecessor media were imported to it, and the techniques and structure peculiar to it took some years to develop; the character of its units was not at once discovered; and credit to a work's creators was for a time thought unimportant. Films had

one special advantage: a standardized technical context existed to make the exchange of units simple.

INTRODUCTION

What do we want to get out of an information system? There is a great need to turn the new machines—computers, mass memories, and dynamic displays—into information systems, or, to be more precise, information storage and supply systems (as distinct from systems for information processing.) But how are they to work, and what are they to do?

What do we want to get out of an information system? The way we answer this rhetorical question probably settles the major aspects of the system we design. The hard-nosed answer is "Documents." The dreamy-eyed answer is "Pure information." The hesitant answer is to ask another question, "What sorts of things is the system for?" Let us probe this sub-question neutrally. If we try to think in terms that avoid the stereotypes and commitments of information retrieval, perhaps we can find a new way of seeing the problem context, and better understand the conventional solutions for what they are and do.

What do we want to get out of an information system? The information problem I will confront here is the overall question of how to keep specialists informed and updated. To give it an acronym, I would venture that existing systems are *Specialist Access and Muster of Information* (SAMI for short). To speak of "access and muster" avoids the connotations of "retrieval"—that the information has been lost, prodigal, or shot down in the bushes. (We will exclude other possible meanings of the term "information retrieval," considering library inventory systems to be a temporary expedient, and management information to be a job for real-time system programming.) I will propose that SAMI is best served, not by document retrieval or content retrieval, each in its way more conventional than it looks, but by something else again. This paper is an elaborate conjecture which I hope to give verisimilitude by an extended historical analogy.

"INFORMATION RETRIEVAL"

Information retrieval in the present sense now means two things. One, document retrieval, uses standardized indexing techniques to draw desired classes of unit documents from a central pool. The other, content retrieval, holds a body of integrated knowledge to which the user may address questions. (For a unified treatment of the two types, see Licklider.[1]) Each is important.

A. DOCUMENT RETRIEVAL

Let us examine the first meaning of information retrieval, document retrieval (DR), the finding of documents having certain indexed descriptions. In document retrieval we index a collection of documents by particular terms or aspects, and later only get things back *through* those terms. This approach retains the unit of the discrete "document," taking it as given. The documents themselves are not expected to have any provision for the system, or to have been created with any relation to the system in mind, but are produced by the world according to external causes. The document is integrally sacrosanct, its content and form untouched by the system. No tampering, excerpting or post-editing of the documents is permissible, no matter how useful it might be.

The problem of the system then becomes finding "the right" documents, which the user recognizes by criteria of his own. The match between the user's private criteria and the available indexing is indicated by the number of "hits"—documents the user is pleased with.

Studies of library use have indicated that not very much library use could be aided by conventional document retrieval systems as currently construed. This would be all right if we could be sure they were good for something else.

That is one criticism of document retrieval: it is not yet clear that document-retrieval systems will do what the users need to have done, even in some ordinary sense, and before we start reforming the users.

The second criticism of document retrieval is based on the nature of the indexing, or categorization, process. (Indexing is categorization, even when it becomes complicated with the relational tagging of roles and links.)

The best categories, and indices, decay. In principle any indexing categories will progressively decline in relevance. It is quite reasonable to expect that as emphasis and interests change in a field, many or most indexing categories for information retrieval will gradually fall apart. A cursory examination of the Dewey Decimal System should confirm this point. If the structure of a particular field strays from the way the original indexing saw it, we will look in the wrong place for the information that was put away earlier.

B. CONTENT RETRIEVAL

The second sense of information retrieval now being used is content retrieval (CR). Content retrieval is often called "fact retrieval"; a system for content retrieval is sometimes called the QAS, or question-answering system. (An excellent source on CR is Simmons.[2])

Such systems can be seen as having two parts; some overall method of storing a complex network of interrelated information, such as that of Marill,[3] and a program mechanism for following through its different parts to find something out. Network and query programs are intertwined; the design of one has consequences for the other. However, for the present purposes we may call such a system the "queriable information network," since the information network is the basic part; in principle a way can presumably be found to interrogate any adequate logistic system.

Content retrieval discards the unit of the document. The total contents of a queriable information network need not be divided into units at all, except those described in the information and those required to represent it.

Such systems offer great promise in a number of ways. However, they do have present and future drawbacks. Whether such systems can replace significant amounts of library use cannot be ascertained. Second, preparing information to go into the queriable information network may be an extremely difficult, touchy and expensive business, at least as things look now. For some systems, coders may require training in mathematical notations of great complexity. Third, conversing with such a machine will on some systems usually, and in the most lucid systems sometimes,[4] require formulating questions in arcane languages or dialects. Either these can be known only to specialists, or their use by the laity may lead to mistakes.

The major problem, though, and the real long-term drawback, is analogous to the decay of categories and indices in DR. Content retrieval assumes that all the information in the network is *consistent and true*, and has no inconvenient gaps. Like any overview, CR must ignore uncertainties and lacunae, and (as "knowledge" is revised and interests shift) it may leave only the tintype of a certain moment's concerns. As scientific and other beliefs are constantly changing, "knowledge" must frequently be readjusted or rearranged.

Actually, the idea of "knowledge" is misleading in this context. What we regard as knowledge includes various contradictory sets of interwoven propositions, clusters of argument, and discrepancies of faith, taste and emphasis in the literature. There is usually disagreement in every field of knowledge. This means that, in a queriable information network, different and mutually contradictory theories really ought to be able to coexist within the system.

Moreover, we may surmise that in most fields of knowledge some non-controversial information is incorrect, but we don't know which. It follows that unless the information network can be modified as systems of information are modified, the information network will be left behind. While evolutionary and patching modifications might be arranged to cope with the

representation of changing views, this expands the problem significantly. (For more on such evolutionary issues and difficulties, see Nelson.[5])

MANIFESTO

I would like to make the claim that document and content retrieval systems are not the principal nor the basic use of computers for handling ideas. I am not claiming that such systems are useless or improper. Rather, I think they will be helpful accessories to the proper, natural and true use of computers for handling the written word.

Systems thinkers have proposed a great variety of possible arrangements for ordering man's thoughts and aiding his mind with the computer. I would like to propose that this diversity of possible facilities and functions is in fact comprisable into an *organized text medium,* to be created and used by people and for people. This organized medium will be that of the text structure created expressly for on-line display by the computer. Such a text will be read from an illuminated screen, the cathode-ray display; it will respond or branch upon actions by the user. It will be a succession of displays, largely textual, that come and go according to his actions. Such a congeries of interconnected text may incorporate and have available many chunks of literature on a given subject, or many subjects, purposely assembled and woven together by authors and editors. The name for this new medium is hypertext.[6]

Hypertext is the combination of natural-language text with the computer's capacities for interactive, branching or dynamic display, *when explicitly used as a medium.* Or, to define it more broadly, "hypertext" is the generic term for any text which cannot be printed (or printed conveniently) on a conventional page, or used conveniently when bound between conventional covers. "Non-linear text" might be a fair approximation. Hypertext may differ from ordinary text in its sequencing (it may branch into trees and networks), its organization (it may have multiple levels of summary and detail), its mode of presentation (it may contain moving or manipulable illustrations, moving or flashing typography), and so on. At last, with the appearance of the computer display and appropriate back-up systems, such texts are possible and practical.

We will store natural-language texts in digital devices for dynamic display, and arrange them into new kinds of units. This means discarding customary document boundaries, and organizing the materials in a complex and non-linear fashion. The user will read from a screen, and as he goes he will make choices that call connected materials into view. Hypertext will be a medium, not a facility, and the institutions of authorship, citation,

academic commentary and argument will be preserved and incorporated in an orderly fashion.

Thousands (or millions) of paragraphs of source material may be available directly to the inquiring reader who wishes further detail, or a certain person's comment, or a specific idea or sentence. Moreover, the *original context* of such items may be summoned, and such contexts—articles, books, or other hypertext units—themselves examined. Thus the reader in search of understanding may continually uncover new elements of interest, often of growing relevance to his hunches, his confusions and his unphrased questions. He can find explanations, ideas, or whole new areas that are to his interest. He may study them through further excerpts, reviews, summaries, or even conventional reading.

The reader will explore, light pen in hand, browsing or studying closely, choosing paths he prefers, criss-crossing these paths on summary levels, stepping from section to section and stopping for deeper work in units of interest. He will come to understand the whole, or a general section, like a walker exploring a city.

Hypertexts will require no special training to use, and rather little special training to create, though like works in other media they can be made either well or poorly. To use, or read, a hypertext, you will need mainly to sit at a screen, read, and indicate choices by pushbutton or light pen. To create a hypertext, you will type in new text and indicate the new relational connections (and their operational meaning) to the system. Where you wish to establish fixed sequences or give a test, you will also state what user's action is to evoke what presentation.

There is reason to suppose that hypertext systems, once created, will serve a range of purposes. These will include the informing of professionals, education, briefings, giving directions, and even providing leisure reading. In the present SAMI context, this generality is not accidental but necessary. While we will not discuss hypertexts for education in this article, we will discuss the need for this generality in satisfying the customary purposes of information retrieval.

Of course it is risky to speculate on what the particular forms and uses of hypertext will be, since its actual presence and demonstration should swiftly outstrip speculation; until then we can have little feel for it. For any but conventional text, we have little notion of the comparative sense or utility of different sequences and structures. All that we can tell at this time is that there are many possibilities. Until the money is spent to try it, speculation is all we can do. But let us do that.

GENERAL TEXT STRUCTURES

Let us consider in the abstract some of the ways that hypertext may be organized. We will assume the use of some machine capable of presenting the texts in any arrangement, sequence or response network that we wish to plan. What ways are there? (To save space and accentuate continuity with the present, we will refer only to "text structures," recognizing that arrangements which depart from the printable collectively deserve the term "hypertext.")

A. TEXT UNITS

Text structures are composed of units. These units may be for convenience(the page of typescript) or have some meaning (the paragraph; the book.) The units may be composed of other units (as chapters are made out of paragraphs) and may be combined to form larger units (as chapters make up a book).

It is important to note that we may discard conventional document boundaries without discarding the notion of unit organization; organized units of text need not have linear sequence, as they do usually. Moreover, organizing text into units need not restrict the types of interconnection among these units that may be tried. We may indeed define unit boundaries in hypertexts, to give creators and users a sense of orientation, and to facilitate administration.

B. TEXT TYPES

Let us then consider different types of texts. Acknowledging that this taxonomy is quite tentative and uncertain, I will try to enumerate some characteristics that texts of all kinds can have. The object is to learn what traits will distinguish the different kinds of text (especially, hypertext) from one another.

Collection vs. Composition. We may distinguish text structures according to the way in which their units are created. We may distinguish between *collected* texts, which are assembled by some method or rule, and *composed* texts, which are put together with consideration of some of their overall properties by some human agency. As examples of collected text we may take a full mail box, a library, or an issue of *The New York Times;* as examples of composed texts we may consider most books or poems, or the front page of *The New York Times.*

Sequentiality. The next obvious distinction is that of sequentiality. The texts may be *sequential,* that is, have a specific order of presentation, or

structural, with the reader free to move among the parts as he sees fit. (Books as we know them are of both kinds, and there are mixed cases as well; again, the newspaper.)

Connective Structure. The connective relations of texts are yet another basis for distinguishing among them. There may be many different patterns of discrete connections among minimal text segments. Similarly, between the larger text units—which may contain many pieces—there may be many degrees of overlap, interconnectedness and penetration. Print technology has limited our experience of text interconnection, but there are simple conventional examples. These include parallel texts (commentaries, parodies) and hooked-on materials (the footnote, the illustration).

Hierarchical relations may also exist among text structures. That is, some units may be "over" others. By convention this hierarchical relation has to do with conceptual generality in the higher unit. However, hierarchy might take on other meanings. Even for the existing "generality" convention, lattice or other connective structures are possible, where the "more general" units could be for different contexts and settings.

The hypertext will have other attributes, based neither on units nor connectedness. These we may deem (depending on various things) to be "at the user interface," at the option of the text's creator, or left to the decisions inherent in an administrative setup. They include the user's freedom to read or use anything contained in the text complex; his ability to take notes, trace his wanderings, or otherwise to couple the use of the hypertext with his own library system; and permission to contribute to the text.

The supporting facility that presents the hypertext must also have—besides displays and calls, indexing, and the ability to cross-follow from one part to another—such other functions as word searches and counts. (However, note that the provision of service functions that speed up lookups and cross-referencing should not be confused with the hypertext medium.) It may also usefully include such complex indexing and queriable information structures as are now associated with DR and CR, just as books now contain indexes.

SUPPLANTING IR WITH HYPERTEXT

The problem is to answer questions and update specialists; that is, to facilitate specialist access and muster of information (SAMI).

Let us consider here a particular text structure for SAMI purposes, the specialist (or scholarship) hypertext. Let me offer speculations about what the specialist hypertext will be like.

The hypertext for these purposes will probably be a particular kind of

anthology. It will be a repository hypertext, continually enlarged and intermittently edited. Its constituent materials will be from time to time assembled into composed units; for instance, a collection of source texts in a given area of discourse. This collection will be connected with many links to specially-composed review articles, arranged in various strata of summary, detail and commentary. By reading in one of these master texts, and turning to original materials or other instructional sections, the user may keep abreast of a field or instruct himself in an unfamiliar area.

The creation and maintenance of these specialist hypertexts will be a cyclical process. First the "original" units or documents will be put in; these will correspond to reports and articles as they are now conventionally published. So assembled, they will constitute a collected hypertext, structured only by minor indexing—cross-referenced, say, by chronology, authorship and topic. These source materials will remain in the system.

From time to time these source materials will be assembled into composed units; without, however, losing anything in structure or independence, since they will continue to be separately addressable. The new composed units will be created by author-editors, like anthologies and review articles. These author-editors will write overviews, summaries and other passages of their own, create connections, make excerpts and digest versions, and add clarifications, arguments, comparisons, diagrams and miscellany.

The units will be signed, like review articles, and thus linked to the professional prestige of their authors. Short squibs and even connective relations may be attributable in this way to particular persons. The authorship will consist of both writing new material and making connections and arrangements among parts of what existed before. It follows that readers who know something may *contribute* materials. This requires an appropriate administrative context. Some people might need to have their work reviewed, others not.

Such a corpus will have new research results, discussion, and interjections continually added to it by identified individuals. It will be continually enlarged, restructured, reconstructed. Periodically the overview-articles will be updated or replaced, with the old ones kept available for the record. Materials of fading interest will not be destroyed; it will just take longer to get to them, and the newcomer will hear of them less.

Dissents may be incorporated. The system imposes no assumption of completeness or infallibility on its constituent parts. Inconsistencies will be welcome; and with all sides of any argument immediately available, the totality of interrelations among divergent viewpoints will be sooner understood.

Such an approach reflects the real structure of scientific and academic

fields. The text network of current interest will grow and change, rather than, in its fixity, depreciating in relevance as a remembrance of the way thinking went at a particular time.

REMARKS

Curiously, the common view of the information system appears to take it for granted that mechanization will mean depersonalization. "Naturally, such a machine will not be popular among authors since by and large no reference will be made to the originator of any given idea. But, on the other hand, perhaps this will be all to the good since it will avoid the hard feelings which are so often produced at the present time by inadequate referencing."[7] There is no need for such dehumanizing; it will work better the other way. The hypertext approach ties in with existing human institutions and motivations. We must discard the idea that efficient mechanization means arbitrary truncation.

It may be complained that we should eliminate the vagaries and biases that individuals composing hypertexts will introduce. To this I would reply that vagaries and biases which are signed by individuals, and for which they are responsible, are far preferable to those that are silently designed right into the system's heart.

To be useful, the hypertext medium requires some degree of prose discursiveness in the material, or variety in the ways and sequences that the same material can be connected together or explained. Indeed, these texts may be made big and diverse enough for study by specialist and beginner alike, with many entrances, tracks and specially-oriented meanders. Thus the user's previous background and level of knowledge could be taken beneficially into account by the author-editors.

It might be claimed that facilities of such a general character are not necessary for the purposes conventionally served by information retrieval. The purpose of retrieval systems is to inform the technical user of what he needs to know, giving short replies to discrete specific questions.

This may sometimes be the case. When technical queries seek only to fill the blanks in some frozen context, DR or CR will do. Often the user has all the background he needs, and no long explanation is needed as an explanatory frame for the short reply. But even the specialist cannot always ask specific questions or understand short answers, let alone cryptic titles and summaries of borderline literature. For such needs these systems really ought to be ready to teach in the larger sense, even when talking to professionals. SAMI systems need the backup capacity for general instruction or review. And so the function of the hypertext, even in the SAMI context,

will be somewhere between "teaching" and "giving information." (How often there seems no feasible practical distinction between these two.)

Such systems will not, of course, replace all information retrieval systems. Where a fixed-context retrieval structure safely tells the user all he needs to know, as does a train schedule, nothing further is required.

UNDERSTANDING THE MEDIUM

Computer branching operations, text storage and display determine this medium. If this is the case, why is the medium taking so long to be developed? I claim that our problem has been not knowing, in some important sense, what the things are *for*, even though these devices have been with us for some fifteen years.

If complex text display and handling is (or determines) a medium, why doesn't everybody know it?

There are two answers to this, which complement each other.

The first is that various people *do* know it, *have* grasped the medium's character. The second is that there is no reason to expect everybody to catch on to a new medium right away. Let me discuss both of these assertions.

To predict that we will have non-linear text media is not very original. Similar ideas have been expressed by many people, beginning with Vannevar Bush,[8] if we do not mention the Oz books and other immemorial sources. Marshall McLuhan has said it in a very general way: this "mosaic" medium is another step in the transition to organic thinking, learning and society that he predicts.[9] The suggestion here encompasses views expressed by Harold Wooster,[10] H. Bohnert and M. Kochen,[11] T. M. Williams,[12] and (in different places) by several contributors to the Intrex Planning Conference.[13] Bush's prediction was the most thorough. But someone identified only as "Our Special Correspondent" also made the point rather precisely in the *Times Literary Supplement*, May 4, 1962, when he (?) said the scholar of the future would "do more than simply manage linear prose. . . ." Rather, he would "cut it like a film or compose it like a picture. . . ."[14]

CULTURAL LAG

Sociologically, we may regard information retrieval (DR,CR) as a somewhat conservative response by the professional community to various technological innovations. It is a time-honored sociological theory (advocated chiefly by W. F. Ogburn[15]) that practices take time to catch up to inventions. This is called "cultural lag," the delay between a change in society (especially a technological change) and the culture's catching up by squar-

ing related practices with the change. Confusion and inefficiency follow the change until the society finds appropriate ways to deal with it.

For reasons both personal and corporate, the irruption of technology in information handling could not be ignored. It had to be assimilated. "Information retrieval" continued the things that had been done before, extending the indexing and abstraction processes. Document retrieval is an advanced card file; queriable information networks are advanced reference charts.

THE MOTION PICTURE ANALOGY

I would like to draw an analogy between hypertext and the motion picture. By considering the time and trouble it recently took to understand the nature of the motion picture—another prepared temporal and visual medium—we may see better why things are difficult now.

Motion pictures are of interest to us in several contexts; in the development of their elements and techniques; in the development of organized units, especially units which are effective and viable; and in the development of the signatory property, the ability to be signed. These comparisons are suggestive analogies rather than close parallels. Let me recite some highlights from the history of motion pictures, so that the sorts of development that occurred, and what was on the minds of its developers, can enlighten our present. (This will concentrate on the photoplay; let it be understood that other uses of the film are derivative.)

The technology of the motion picture was not suddenly created in the shop of any one inventor. However, Edison's invention, first tried with film in 1889, was the one that took. He called it the kinematograph. It permitted fast viewing of successive frames on a long strip of film. Edison did not work on a projector because he and his advisors thought it would destroy the novelty. He also failed to patent his machine abroad. In 1895 Thomas Armat made a working projector for Edison's film. About the same time, the brothers Lumière, photographic manufacturers in France, created an all-purpose movie machine: the Cinématographe, which could be used by turns as camera, printer and projector. They promptly took it on the road as a novelty, showing films of local scenes, the arrival of trains, and the marching of troops.

The next year, 1896, a real showman began making movies in France. His name was Georges Méliès. A magician and theater operator, he at once saw the potentiality of the medium for presenting the astonishing and fantastic. He made many science-fiction and magic shorts. Méliès originated fades, dissolves, animation, and the whole range of trick photography and

multiple exposures. He is also credited with the invention of editing. With *Cinderella* in 1900 Méliès is said to have produced the first story film.[16]

Meanwhile, films in America were already popular, having been introduced to vaudeville in 1896. Credit for being the first "theatrical film" is usually given to Edwin S. Porter's *The Great Train Robbery* (1903). However, Porter's *The Life of an American Fireman*, produced the year before, had contained all the same elements.

A curious monopoly circumstance, with interesting consequences, gradually came about in the American motion picture industry. Because various companies had tried to circumvent the Edison movie patents, there were many lawsuits over the process around the turn of the century. The legal tangle was cleared up by embracing the contending film companies into a monopoly. This monopoly, the General Film Trust, controlled all film production and distribution in America under the umbrella of the Edison patents. For a brief time (1909 to 1915) this trust produced films like sausages, undifferentiated, with the names of the makers and actors not revealed. The officers of this trust preferred that the works they produced be marketable as an interchangeable commodity of a single length (one reel) and quality (mediocre).

D. W. Griffith, the founder of the motion picture art and perhaps the greatest film director of all time, began directing anonymously during this period. In his early work with the Biograph Company, beginning in 1908, he experimented in every possible way with the construction of scenes and narratives, and soon surpassed every other director. Between 1908 and 1915 he invented and consolidated modern camera technique, modern editing technique, and modern narrative technique. By the end of this period his name was known. It was he that consciously began constructing the film out of *shots* rather than "scenes." Griffith's *The Birth of a Nation*, released in 1915, was a "total" film the way none had ever been. Uniting all the techniques that had come before, all his new effects of camera placement, cross-cutting, and tempo of movement and editing, he wove history, battles and love story into a single, entire fabric. Audiences were stupefied. Woodrow Wilson and Lenin were both deeply moved by the film.

Thus the twenty years from 1895 (Armat and Lumière machines) to 1915 (*The Birth of a Nation*) saw the development of a complete medium, in its scope, techniques, and units. No one caught onto it overnight. Nobody saw at the outset what it was going to become.

Now let us examine selected aspects of this brief history. The correspondences to the present situation may be evident.

A. THE MEDIUM'S MISUNDERSTOOD CHARACTER

If we consider the way some of the early things were done, it is by no means obvious that people working in the film at that time had the faintest idea of what was to come. Indeed, it seems that at least some of the film's originators did not know what they were doing or had done. Edison did not want to develop a projection system for his Kinematograph because he thought the novelty would wear off. And when Méliès asked to buy the Cinématographe, Auguste Lumière told him: "It can be exploited for a certain time as a scientific curiosity, but, apart from that, it has no commercial future whatsoever."[17]

Through a misunderstanding of the medium, and a failure to recognize its character, film-makers repeatedly imported to it the techniques of the stage. While stage *content*, when adapted, has worked, stage techniques have not worked in films, and have given satisfaction only for short periods of time.

The use of flat backdrops, action in a plane facing the camera, and the continuous take throughout a whole scene were all early conventions abstracted from the stage. The dramatic scene of early films was called the "action tableau," but it had relatively little action; its flat arrangement was a simple transposition from the stage, and its slight motion a prudent simplification of ordinary staging. This was the technique of Méliès, who also used painted backdrops shamelessly. Though the action tableaux, after Porter, became infused with more and more action, heavy stage influence remained. Even Porter was not completely won over by his own inventions. Though he had pioneered three-dimensional action and stirring visual effects in *The Great Train Robbery*, when he filmed *Uncle Tom's Cabin* he reverted to tableaux, flat sets and flat action.[18]

Stage plays have often been filmed as is; this occasionally happens even in the present day. It does not work very well. In 1907 the Film d'Art Company, in France, was established to uplift the motion picture by filming plays. Their product, while hailed for being cultured, was not very good. In 1912 Adolph Zukor's *Queen Elizabeth* opened. It was an attempt to immortalize Sarah Bernhardt on film. It made money. It also showed rather clearly that the sorts of posture, makeup, movement and declamatory style thought appropriate for the stage were not at home in the new medium.

But Griffith clearly saw the differences between stage and film. He was asked whether he thought knowledge of the stage was necessary to direct films. He said,

"No, I do not . . . The stage is a development of centuries, based on certain fixed conditions and within prescribed limits. It is needless to point out

what these are. The moving picture, although a growth of only a few years, is boundless in its scope and endless in its possibilities. . . . The conditions equally dissimilar. . . ."[19]

B. TECHNIQUES

There is a basic roster of techniques of film. These are effects (stop-motion, animation), display arrangements (camera angles and movement), and the selection and interconnection of parts (direction, editing).

These techniques of film are basic and perfectly known to those who make films, though there is disagreement on style and appropriateness. All the techniques are familiar. Nothing is new, in the sense that it is impossible today to create new elements. There has been virtually nothing new in the last 50 (fifty!) years, since, say, Griffith's *Intolerance* (1916).* But these techniques of film were by no means obvious. It took twenty years of film-making to discover them, from 1895 to 1915. Though many of the films have been lost, in principle each innovation has associated with it a specific film, director and date. The various techniques were introduced one by one, chiefly by Méliès, Porter and Griffith.

Early movie-makers did not use more than one shot in a scene. They did not move the camera. They instructed their actors to wave their hands around and exaggerate every gesture, since the camera was so far away. And the camera was so far away because film-makers thought it looked wrong not to show the actor's whole body. Other rules included:

"1. Every scene must begin with an entrance and end with an exit, just as on the stage.

"2. Players must face the camera and move horizontally, except when the movement was rapid, as in a chase, or prolonged, as in a fight. At these times action was in diagonal relation to the camera in order to give the players more area.

"3. Any action in the background must be slow and greatly exaggerated so that it would 'register' on the audience.

"4. Pantomime must be exaggerated and over-deliberate; e.g., a stare had to be held; a start must be violent; speeches must be mouthed with pronounced slowness."[20]

C. UNITS AND THEIR ORGANIZATION

The character of basic units, in particular the narrative film, was not at

---

*Sound, color, etc., did not modify these techniques; but they added separate problems of their own and additional opportunities for synthesis.

first grasped. "In 1896, a special outdoor performance by Joseph Jefferson of *Rip Van Winkle* amidst the natural scenery of his summer home on Buzzard's Bay, Massachusetts, was recorded by Biograph on 219 feet of film. But instead of shooting this subject with a view toward an interesting and a dramatic continuity, it was filmed in eight episodes billed and sold as separate catalogue numbers. . . ."[21]

For a while, projected films were timed to the vaudeville turn, and then to the organization and taste of the nickelodeon theater. Though films tended naturally to become longer, after about 1908 The General Film Trust held the length of all Trust films to one reel (1000 feet, about sixteen minutes). It was not until the decline of the Trust that the American film could break out of this one-reel mold and attain feature length.

Not only was the overall unit misunderstood; the parts themselves were not thought to be manipulable and discrete. The basic unit was thought to be the "scene." It was only D. W. Griffith who learned to work in terms of *shots* rather than complete scenes; and to construct the scenes from the ground up, using the shot as the building block. Griffith discarded the principle that a scene must end before a new shot may begin, and, indeed, the principle that one scene must end before another begin.[22] By manipulating the constituent parts he learned to handle the motion picture as a whole, controlling and integrating all its aspects.

The climax, of course, was Griffith's *The Birth of a Nation*. Though panoramic spectacle with three-dimensional sets had been used before, especially in the Italian *Quo Vadis*, it had not previously been integrated with a well-told story. *The Birth of a Nation* was such a complete, unified, spectacular achievement: the true feature film, and a model seldom equalled.

D. THE SIGNATORY PRINCIPLE

"Films are made by men, not machines or corporations."[23] Yet interestingly enough, the notion that films should be *signed*—that is, have personal credit go to their creators—was not acknowledged for a time. Quite the contrary: credit was for a time *suppressed*.

The film trust's management of the motion picture industry from 1909 to 1915 produced an even stream of one-reel films whose actors and directors were not known to the public. The public was developing favorites, and demanding their names—like that of the girl everyone knew as "Little Mary." The exhibitors, too, wanted the names so they could in some measure control public response, i.e., book the favorites and advertise. But the favorites' names were kept a secret. (The film trust's policy of not releasing the names of actors was ably met abroad by the simple expedient of assigning names to them.[24])

The cat got out of the bag in 1910. Carl Laemmle of the trust-defying Imp Company publicly advertised the name of a film actress for the first time, after he had stolen her from a trust studio. Thus began the star system. Competition from Laemmle and other outlaw producers forced the trust companies to reveal names. ("Little Mary" found it expedient to adopt the name of Mary Pickford.) Griffith's name was first advertised when Biograph gave in to the new open-names policy in 1913.[25]

IMMANENCE

I have attempted to point out some curious similarites between a general-purpose medium we all know and a general-purpose medium that none of us, to my knowledge, has seen.

The comparison would be strongest, of course, if we based it on a premise of structural determinism. This would be to assert that media like the hypertext and film derive their character from intrinsic principles and internal dynamics that are somehow built right into them.

This was the view held by Kuleshov, the great Russian teacher of film. He said: "In every art there must be first a material, and secondly, a method of composing this material specially adapted to this art."[26] As the reader sees, it took several decades to discover these for the film.

Perhaps in general we have no leeway, then, and media are ours merely to discover rather than to invent. But I am not sure that for the hypertext it is this tight. We can in hypertexts establish precedents and traditions, doctrines and schools; we can, by stress upon programmed instruction, flash card effects or civil liberties, influence for worse or better its usefulness to our society and condition.

The medium has an internal dynamic, I think, immanent and intrinsic in the technology. This we must discover. But however compelling this internal dynamic may be, there will be play for some adjustments. We must strive to add the touches that suit it best to the way we want to do things.

A DISANALOGY: STANDARDIZATION

By luck, movies are standard. The width and sprocketing of Edison's original films became and remained the 35mm motion picture standard. While there are other standards for other purposes (8mm, 16mm) and other rival processes (3-D, introduced in 1922, and such other intermit-

tent specials as wide-screen and smell track), in general, a movie is a movie.*

In order to try out the new medium, other than haphazardly, standardized settings must be created, so that work in this medium can be used and passed around. We need standardized character-sets and connector notation, standard ways of scaling down a presentation to a sparse facility and of interfacing it through varying screens, buffers and accessories. This might be accomplished by the specification of standard facility levels, so the creators of such things could either stipulate a minimum system required for use of a given text, or an appropriate presentation level for each configuration of hardware.

Moreover, it will be necessary to establish reliable royalty provisions for authors, editors and contributors, probably based on the presentation frame. But something equivalent to *ownership of a copy* will also be necessary, so that a given user may have unlimited access to his own possessed material.

CONCLUSION

Information retrieval has extended traditional techniques to new technology. An analogy has been found in the history of motion pictures. The prior techniques of the theater were at first extended to movies, and it was several decades before we understood the motion picture as a medium. It was as natural to extrapolate these prior techniques to the movie as it now seems natural to extrapolate prior techniques of information handling to the CRT-computer-mass memory system. Here again, as in films, the technology makes possible a medium which is radically new and has dynamics of its own. The most suitable techniques and units were not immediately evident in movies, either.

The conventional "document" is not God-given, and in fact is inappropriate for most purposes. Systems based on discrete and isolated documents relinquish the greatest power of the new technology. But at the other extreme we are wrong to suppose that an information machine can or should eliminate the human task of composition. We have not begun to explore the possibilities of natural language woven into more complex (but also natural) arrangements.

The problem of Getting It Out of Our System, then, is not the problem

---

*It is important to note that facilities are readily available for the transposition of off-standard work to the main standard, by optical printing and sound transfer. Often novelty effects offered in first-run theaters are scaled down to the standard for distribution.

of fishhook design for a document pool, or of creating a conversational black box with a narrow vocabulary. We must get out of our system the fixities of thinking and procedure that hold us back.

CITATIONS

1. J. C. R. Licklider, *Libraries of the Future*. M.I.T. Press, 1965.

2. R. F. Simmons, "Answering English Questions by Computer," *Communications of the ACM 8:1* (January 1965), 53-69.

3. Thomas Marill, "Libraries and Question-Answering Systems." Report Number 1071, Bolt Beranek and Newman, Inc.

4. Bert F. Green, Alice Wolf, Carol Chomsky & Kenneth Laughery, "Baseball: an Automatic Question-Answerer" in E. A. Feigenbaum and J. Feldman (eds.), *Computers and Thought*. McGraw-Hill, 1963, 207-216.

5. Theodor H. Nelson, "A File Structure for the Complex, the Changing and the Indeterminate." *Association for Computing Machinery Proceedings of the 20th National Conference*, 1965, pp. 84-100.

6. Theodor H. Nelson, "The Hypertext." Paper presented at 1965 Congress of the International Federation for Documentation, Washington, D.C., October 1965.

7. Andrew W. Booth, "The Kinds of Machine Now in Use" (No. 4 in the series "Freeing the Mind"), *The Times Literary Supplement*, April 13, 1962, p. 252.

8. Vannevar Bush, "As We May Think." *Atlantic Monthly* 176:1 (July, 1945), 101-108.

9. Marshall McLuhan, *Understanding Media: the Extensions of Man*. McGraw-Hill, 1965.

10. Harold Wooster, "A Web of Words." Keynote Address of Sixth Annual Institute in Technical and Industrial Communications, Colorado State University, Fort Collins, Colorado, July 1963.

11. H. Bohnert and M. Kochen, "The Automated Multilevel Encyclopedia as a New Mode of Scientific Communication," in Manfred Kochen, *Some Problems in Information Science*. Scarecrow Press, 1965, 156-160.

12. T. M. Williams, "Some Logical Aspects of Thematic Representation with Explicit Ties." Paper presented at 1965 Congress of the International Federation for Documentation, Washington, D.C., October 1965.

13. Carl F. J. Overhage and R. Joyce Harmon, *Intrex: Report of a Planning Conference on Information Transfer Experiments*. M.I.T. Press,

1965. Two notes by J. C. R. Licklider are of special interest: "An On-Line Information Network," 147-155, and "Proposed Experiments in Browsing," 187-197. See also E. S. Proskauer, "The Motivations of Authors—Intellectual Property and the Computer," 199-201.

14. Our Special Correspondent, "Poetry, Prose and the Machine" (No. 7 in the series "Freeing the Mind"), *The Times Literary Supplement*, May 4, 1962, p. 310.

15. William F. Ogburn, "Cultural Lag as Theory." *Sociology and Social Research* 41, 167-174. Jan-Feb 1957.

16. Richard Griffith and Arthur Mayer, *The Movies,* Bonanza, 1957, p. 6.

17. Maurice Bardèche and Robert Brasillach, *The History of Motion Pictures* (tr., ed. Iris Barry), W. W. Norton Co. and the Museum of Modern Art, 1938, p. 10.

18. A. Nicholas Vardac, *Stage to Screen: Theatrical Method from Garrick to Griffith.* Harvard University Press, 1949, p. 171.

19. Lewis Jacobs, *The Rise of the American Film: a Critical History.* Harcourt, Brace, 1939, pp. 118-119.

20. Jacobs, *op. cit.,* p. 60.

21. Vardac, *op. cit.,* p. 169.

22. Ernest Lindgren, *The Art of the Film.* London: Allen and Unwin, 1949, p. 71.

23. Iris Barry, quoted by Richard Griffith in "Moving Pictures," *Encyclopedia Americana,* 1960 edition, v. 19, p. 540.

24. Griffith and Mayer, *op. cit.,* p. 47.

25. Jacobs, *op. cit.,* p. 88.

26. This is Pudovkin's reminiscence, cited by Lindgren, *op. cit.,* p. 76.